IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

Detection of Lung Cancer Using Backpropagation Neural Networks and Genetic Algorithm

Mr. Akshay Jadhav[1], Ms. Jennifer D'Cruz[2], Mr. Virendra Chavan[3], Ms. Ashvini Dighe[4], Prof. Jayashree Chaudhari[5]

Department of Computer Engineering, Dr. D.Y. Patil School of Engineering, Pune, India [1,2,3,4,5]

Abstract: Carcinoma also known as Cancer is the commonly growing and most dangerous disease occurred in human species. Lung Carcinoma is one of them. It is a disease that occurs due to uncontrolled growth of cancerous cells in the tissues of the lungs. Prior diagnosis of the disease saves huge number of lives, failing in which may lead to other severe problems causing sudden fatal death. Motive of this system is to automate the detection process so as to perform advanced detection of the disease in its early stage. A measure for early diagnosis mainly includes X-rays, CT images etc. In this system firstly we use techniques such as Data Preprocessing, Training and testing of samples that are necessary for the task of medical image mining. A powerful learning model i.e (BPNN) is used for classification which would classify the digital X-ray, CT-images, MRI's, etc as cancerous or non-cancerous. Further Genetic Algorithm will be used that would extract feature on the basis of the fitness function. The selected feature will help to detect the stage of cancer by measuring the size of the feature and using this measurement the stage of cancer will be decided. This system will support to make an relevant decision about a patient's state.

Index Terms: Backpropagation, Neural Networks, Classification, Genetic Algorithm, Medical Image Mining.

I. INTRODUCTION

One of the leading causes of deaths in both men and women is Lung Cancer. In most of the cases manifestation of lung cancer in a patient's body is divulged through early symptoms. Treatment and prognosis depend on the type of cancer, the stage of the cancer and the patient's performance. Possible treatments include chemotherapy, surgery and radiotherapy. The patients survival depends on the stage of cancer, patients health, recovery and other factors. As currently there are no technical methods to prevent cancer, early detection represents an acute factor in the treatment and increasing the survival rate.

Medical images are a significant part of medical diagnosis and thus for the treatment we concentrate on these images for enhanced results. The medical images include hidden information and features that are utilized by the doctors in making relevant decisions about a patient situation. Extracting and using this relevant hidden information is a critical first step. Hence we need to adopt data mining techniques for knowledge extraction from medical images.

We are using the data mining technique of data preprocessing which includes gray scale and dark image transformation to get an enhanced quality of image for the further process of feature extraction.

The BPNN is being trained and tested using an available database. The features extracted are used for detecting the stage of cancer based on the metrics of the feature used in the system. Thus this system will help the doctors to conclude or make an efficient decision about a patient's condition.

II. LITERATURE SURVEY

There are innumerable existing system that shows various methods of detecting the lung cancer. One such system proposed by Sandeep Kumar Saini[1] propounds the use of fuzzy logic for the testing and accuracy of the features extracted and the ACO technique to augment classification of the cancerous images. In this system Histogram-Equalization is applied to the DICOM images for their enhancement. After the image enhancement, the process of feature extraction is carried out using the Binarization Approach which implicates normal and abnormal images. The system detects lung cancer in its early stages using fuzzy logic and ACO technique. Comparatively Zakaria Suliman Zubi [2] proffers the technique of generalized Neural Networks and Association Rule Mining in order to identify the features of the images and classify them into cancerous and non-cancerous. In this system the image first goes through preprocessing phase. In the later phase the feature extraction process is carried out followed by Rule-generation process and using which the images are classified as cancerous or non-cancerous with the help of Neural Network for the same.

V. Krishnaiah [3] proposes a system which uses techniques such as If-Then Rules, Decision-Tree, Naïve Bayes Classifier, and Artificial Neural Network. The Decision tree is used to drill through the database and interpret it. The Naïve Bayes, IF-THEN rule and Artificial Neural Network are used to classify the database. Thus the results obtained after classification are used by the lung cancer prediction system for early detection which can save a patient's life.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

TABLE 1.1 EXISTING SYSTEMS AND CLASSIFICATION METHODS USED

Sr No.	Paper	Author	Classification Methods used
1	Detection Of Lung Carcinoma using Fuzzy Logic and ACO Techniques	Sandeep Kumar Saini, Gaurav, AmitaChoudhary	Fuzzy Logic and ACO Technique.
2	Using Some Data Mining Techniques For Early Diagnosis Of Lung Cancer:	Zakaria Suliman Zubi and Rema Asheibani Saad	Neural Networks, Association Rule Mining.
3	Diagnosis Of Lung Cancer Prediction System: Using Data Mining Classification Techniques	V. Krishnaiah, Dr. G. Narsimha, Dr. N. Subhash Chandra	If-Then Rules, Decision-Tree, Naive Bayes Classifier, Artificial Neural Network.
4	Identifying Hotspots In Lung Cancer Data Using Association Rule Mining	Ankit Agarwal and Alok Choudhary	Association Rule Mining.
5	Implementation of automatic detection of lung cancer using Adoptive Neuro Fuzzy system	S. Navin Kumar, R. Rishikesh,Dr. M. Thangamani, V.Santhosh kumar, A. Senthil Karthick	Adoptive Neuro Fuzzy System
6	Early Detection of Lung Cancer Risk Using Data Mining	Kawsar Ahmed, Abdullah-Al Emran, TasnubaJesmin, Roushney Fatima Mukti, MdZamilur Rahman, Farzana Ahmed	Apriori Algorithm

Ankit Agarwal [4] proposes a system that performs Association Rule mining analysis on lung cancer data to identify hotspots in the cancer data, to examine the patients' survival time which is significantly higher than or lower than the average survival time. The system uses SEER database which is made from the SEER website on license agreement, the data mining method of association rule mining is implemented to classify the images as cancerous or non-cancerous. A two stage association rule mining is used where the expendable rules from stage 1 are discarded in stage 2 and classification is enforced to identify hotspots in lung cancer data.

For automatic detection of lung cancer, Dr. M. Thangamani [5] proposes a system which uses adoptive neuro fuzzy system. The system extracts the affected region from the lung CT image using morphological reconstruction followed by nodule segmentation which is done using global threshold and morphological operations. In the advanced stage the feature extraction is executed on these segmented nodules and based on these features the classification of cancerous or non-cancerous image is carried out with a support of ANFIS based classifier.

Early Detection of Lung Cancer Risk Using Data Mining the research article by Kawsar Ahmed [6] acknowledges the use of methods like Apriori Algorithm. The article propounds the use of a lung cancer database, applying the preprocessing methods on to the database and then disclosing various patterns from the data using the Apriori algorithm. After the patterns are disclosed a clustering algorithm is used to create clusters having patterns with similar ranges of risk. Thus using these data mining techniques early detection of lung cancer risks in various ranges can be achieved.

III. SYSTEM IMPLEMENTATION

The first step is to collect the Lung CT images to provide them as input for the system. The next step will be to apply Data Preprocessing on the images, for image enhancement in order to get the best quality of images. After enhancement of the image, the training and testing of the Neural Network is done using a training database. Feature extraction of the input images is done using Genetic Algorithm.

According to the features extracted, the input images are classified as cancerous or non-cancerous using BPNN. The feature extracted by using Genetic Algorithm is then used to decide the stage of cancer. On the basis of the metrics of the extracted feature, the stage of cancer is decided by classifying the metrics using BPNN.

The processes used for system implementation are described as follows:

1. Image collection: Images are collected from the different available databases. An individual database is stored for the required work. The images from the database are used for the training and testing of the Neural Network. The lung cancer images are in the format of .dicom or.jpg file.

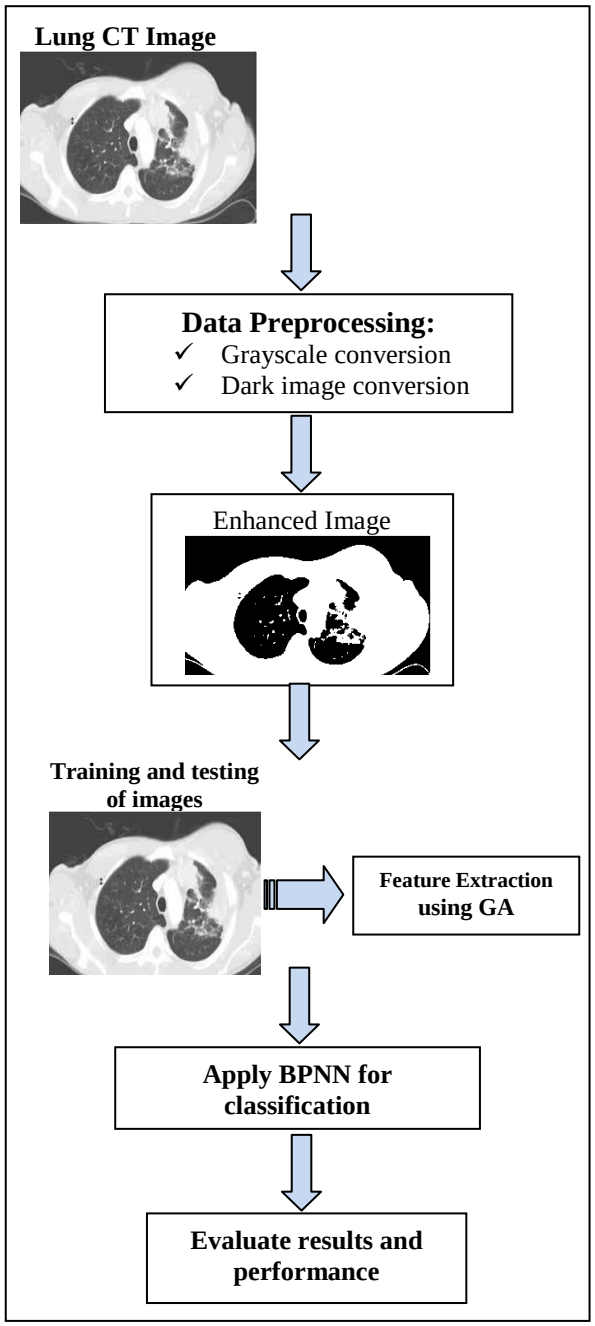
2. Data Preprocessing: Preprocessing is the initial step in the detection of lung cancer. This step is important to improve the quality of the image so as to make feature extraction more reliable. The input RGB images are transformed to Gray-scale and Dark images using algorithms such as Lightness, Average, and Luminosity to obtain its compressed representation.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

IV. SYSTEM ARCHITECTURE



3. Training and testing samples: Training and testing of samples is carried out using the available database. The artificial neural network has to be initially trained with a training dataset for learning and performing classification. Then the artificial neural network is provided with testing dataset .The results of the classification on the test cases are evaluated to check the error frequency or the error rate occurring in the classification. The errors occurred are rectified by changing the weights in the dataset. To start the process of training the initial weights are chosen randomly.

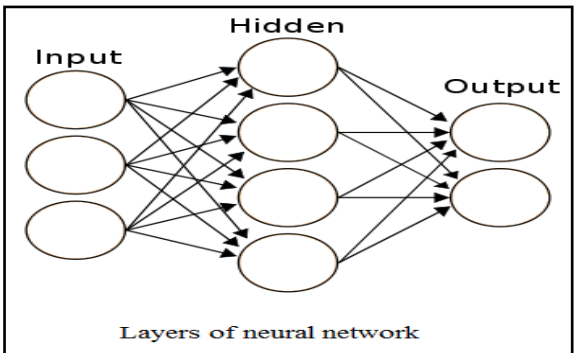
4. Feature Extraction: Medical images have a large number of features and it is important to extract and use the required and essential number of features for minimizing the complexity of processing. In domain of image processing feature extraction is performed on the images for retrieving the values i.e (features) that prove to be informative, non-redundant and help in the learning and knowledge gathering about the given images for carrying out better human understanding and interpretations. Feature extraction can also be called as Dimensionality Reduction. The various features that can be extracted from the image are average gray level, uniformity, standard deviation, skewness, smoothness, contrast, correlation etc. This process of feature extraction is carried out using Genetic Algorithm. The tumor caused due to multiplication of cancer causing cells in the lung is the feature extracted in the implemented system. Genetic algorithm is used for this process of feature extraction in the implemented system.

5. Classification: There are various classification approaches available like fuzzy logic, SVM, neural networks, association rule mining etc. In the proposed system we are making use of the artificial neural networks for classification. We are making use of a classification method i.e. Backpropagation neural network for classifying the images as Cancerous or Non-cancerous on the basis of the features extracted. On the basis of the feature extracted i.e. (tumor) the given input image is classified as Cancerous or Non-cancerous also on the basis of metrics of the tumor i.e size of the tumor, the stage of cancer is classified according to the range of tumor size.

6. Neural Networks: Neural network is an interconnected web of neurons that transmits elaborate patterns of electrical signals. The human brain is a type of biological neural network. An artificial neural network is a cluster of learning models based on biological neural networks.

Advantages of Neural Network:

- Neural networks are compatible and flexible in changing the environment.
- A neural network is capable of learning and recognizing the patterns which exist in the dataset
- Neural networks can easily develop information providing models and handle very complex conditions.



7. Backpropagation Neural Network: BPNN is a learning algorithm used for training the artificial neural network. Mainly, the backpropagation algorithm consists of two stages i.e. - forward pass and backward pass

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

through which the various layers or sections of the network are trained.

The algorithm for BPNN can be given as follows:

1. The first step is to initialize the weights randomly.
2. In second step, an input vector pattern is provided to the network.
3. Check the outputs of the network by directing input signals forward.
4. Calculate $\partial_j = (y_j - d_j)$ for all output neurons, where $d_j$is the desired output of neuron j and yj is its current output: $y_j = g(\sum_i w_{ij}x_i) = (1 + e^{-\sum_i w_{ij}x_i})^{-1}$, assuming a sigmoid activation function.
5. For remaining neurons (from last hidden layer to first), compute $\partial_j = \sum_k w_{jk} g'(x) \partial_k$, where ∂k is the ∂_j of the succeeding layer, and $g'(x) = y_k(1 - y_k)$.
6. Update the weights according to: $w_{ij}(t + 1) = w_{ij}(t) - y_i y_j (1-y_j) \partial_j$, where, w_{ij} is a parameter called the learning rate.
7. Go to step 2 for a some number of iterations, or until the error is decreased to a prespecified value.

8. Genetic Algorithm:

This is a heuristic search algorithm based on the method of natural selection and genetics. It provide solutions for optimization problems using methods such as inheritance, mutation, selection and crossover.

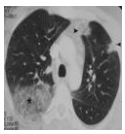
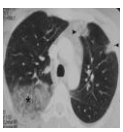
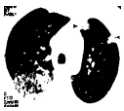
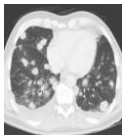
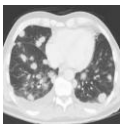
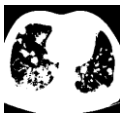
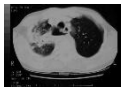
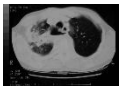
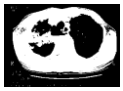
Genetic Algorithm is given as:

1. Generate a random initial population (t).
2. Evaluate a fitness for every individual in the current population.
3. Identify and choose the fittest features i.e. parents from the population.
4. Perform crossover on the parents creating a population (t+1).
5. Now, perform mutation on the population (t+1), which in turn means make appropriate changes in every individual of the population.
6. Discover the fitness of the lastly generated population.
7. Repeat steps 2 to 6 until the situation where the best features meet a pre specified minimum criteria.

Flow Algorithm:

1. START
2. Login
3. Upload the X-ray images or CT-images of the lungs to be tested for lung cancer.
4. The uploaded images then go through the process of Preprocessing, in which it is converted from RGB to Grayscale and dark image.
5. The artificial neural network is trained and tested using the learning database to perform pattern classification.
6. Genetic Algorithm is used to extract features from the enhanced image. The extracted feature is the tumor which is created by multiplication of cancer causing cells in the lungs.
7. The artificial neural network classifies the uploaded images as cancerous or non-cancerous using Backpropagation Algorithm.
8. On the basis of metrics of the feature extracted i.e tumor size, the stage of cancer is displayed as per the range of the tumor.
9. The results are displayed.
10. Logout.
11. STOP.

V. RESULTS

SR. NO.	Uploaded Image	Gray Scale Image	Dark Image	Range of tumor size	Stage of cancer
1				0cm	Non Cancer-ous
2				0.1 cm-5.0 cm	First Stage Cancer
3				5.1 cm-7.0 cm	Second Stage Cancer
4				Larger than 7cm	Third Stage Cancer

VI. FUTURE SCOPE

Future Scope rests in the technique where we can use BFO, Bacterial foraging optimization algorithm (BFOA) which has beenwidely adopted as a global optimization algorithm for distributed optimization and control. The social foraging behaviour ofEscherichia coli inspires the use of BFOA. BFOA has already drawn the attentionof researchers because of itsability in solving real-world optimization problems arising in several application domains. So, the future work suggests hybridization of Genetic Algorithm and Particle Swarm Optimization whichhas the ability to address the above problems.

VII. CONCLUSION

We have studied an approach towards the system for the early detection of lung cancer which will benefit the patient, as the disease can be cured in its early stages and will also help to increase the survival rate. This system provides a framework for detecting the lung cancer using the Lung CT image/X-ray chest image and providing the results concerning a particular patient's health state by making the use of Backpropagation Algorithm and the Genetic Algorithm.

VIII. ACKNOWLEDGEMENT

We would like to thank Dr. D.Y. Patil School of Engineering for providing us with all the required amenities. We would thank our guide Prof. J. L .Chaudhari for giving us all the help and guidance we needed. We are also grateful to Prof. S. S. Das, Head of Computer Engineering Department, DYPSOE, Lohegaon, Pune for their indispensable support, suggestions and motivation during the entire course of the project.

REFERENCES

[1] Sandeep Kumar Saini, Gaurav, Amita Choudhary, Detection Of Lung Carcinoma using fuzzy Logic and ACO Techniques, IJERT, Vol.3 Issue 8, pp.903-906,2014

[2] V. Krishnaiah, Dr. G. Narsimha, Dr. N. Subhash Chandra, 2013, Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques, International Journal Of Computer Scienceand Informayion Technologies,Vol.4(1),2013,39-45.

[3] Zakaria Suliman Zubi and Rema Asheibani Saad," Using Some Data Mining Techniques for Early Diagnosis of Lung Cncer", Rescent researches in Artificial Intelligence, Knowledge Engineering and data Bases, Libya,2007

[4] Ankit Agarwal and Alok Choudhary," Identifying Hotspots in Lung cancer Data Using Association Rule Mining,, IEEE, pp.995-1002, 2011.

[5] J. David Schffer, Angel Janevski and Mark R.Simpson,AGenetic Algorithm Approach for Discovering Diagnostic Pattern in Molecular Measurement Data,,IEEE,2005.

[6] P. Nithya, B. Umamaheshwari, R. Deepa, Detection Of Lung Cancer Using Data Mining Classification Techniques, IJARCSSE, Vol. 5, Issue 7, July 2015.

[7] S .Navin Kumar, R. Rishikesh, Dr. M. Thangamani, V. Santhosh kumar, A. Senthil Karthick, Implementation of automatic detection of lung cancer using Adoptive Neuro Fuzzy system, IJAST.

[8] Tac-sun-choi, "Genetic programming-based feature transform and classification for the automatic detection of pulmonary nodules on computed tomography images", Journal of information science, 2012, Vol. 212, Pp. 57-78.